

# Wavelet-Based Feature Extraction for Handwritten Numerals

Diego Romero, Ana Ruedin, and Leticia Seijas

Departamento de Computación, Facultad de Ciencias Exactas y Naturales,  
Universidad de Buenos Aires  
Pabellón I, Ciudad Universitaria (C1428EGA) Buenos Aires, Argentina  
{dromero, ana.ruedin, lseijas}@dc.uba.ar

**Abstract.** We present a novel preprocessing technique for handwritten numerals recognition, that relies on the extraction of multiscale features to characterize the classes. These features are obtained by means of different continuous wavelet transforms, which behave as scale-dependent bandpass filters, and give information on local orientation of the strokes. First a shape-preserving, smooth and smaller version of the digit is extracted. Second, a complementary feature vector is constructed, that captures certain properties of the digits, such as orientation, gradients and curvature at different scales. The accuracy with which the selected features describe the original digits is assessed with a neural network classifier of the multilayer perceptron (MLP) type. The proposed method gives satisfactory results, regarding the dimensionality reduction as well as the recognition rates on the testing sets of CENPARMI and MNIST databases; the recognition rate being 92.60 % for the CENPARMI database and 98.22 % for the MNIST database.

**Keywords:** Continuous Wavelet Transform, Dimensionality Reduction, Pattern Recognition, Handwritten Numerals.

## 1 Introduction

Automatic recognition of handwritten numerals is a difficult task because of the wide variety of styles, strokes and orientations of digit samples. The subject has many interesting applications, such as automatic recognition of postal codes, recognition of amounts in banking cheques and automatic processing of application forms. Good results in numerals recognition have been obtained with neural networks among other classical learning procedures. The performance of these classifiers strongly depends on the preprocessing step. The choice of features to be extracted from the original data remains a challenge: it is not always clear which ones are to be selected to efficiently characterize their class, and at the same time provide a significant reduction in dimensionality.

Wavelet transforms have proved to be a powerful tool for image analysis, because of their capability to discriminate details at different resolutions. They have given good results in edge detection [1] and texture identification [2]. Discrete wavelet transforms (DWT) have been used to extract features for digit

recognition. A 1D DWT [3] and a 1D undecimated multiwavelet transform [4] have been applied onto the previously extracted contour of the digits, and the result fed into a MLP classifier. Multiresolution techniques have also been used in conjunction with more complex classifiers: after applying a 2D DWT (3 different resolution levels) to the digits, a combination of multiple MLPs was used, each one being trained, with dynamic selection of training samples, for a specific level of (thresholded) approximation coefficients [5].

On the other hand, the 2D Continuous Wavelet Transform (CWT) performs a scale-space analysis on images, by calculating the correlation between an image and a 2D wavelet, at different scales and locations. It is more flexible than the DWT, and allows for scales that are non dyadic, providing greater frequential resolution. The 2D CWT has been extended by giving one principal orientation to the wavelet, via stretching one of its axes, and adding a rotational angle as a parameter to the transform [6]. It has translation, rotation and scale covariance [7] (the CWT is covariant under translations, because when applied to a translated image, it produces the translated CWT of the original image). This CWT has been applied for pattern recognition in images [8], and has given satisfactory results for digit recognition [9]. We use it here to extract a shape-preserving smaller version of the digits and to build a complementary vector with information on orientation, gradients and curvature at different scales.

We implemented the recognition system using a feed-forward neural network trained with the stochastic back-propagation algorithm with adaptive learning parameter. Our experiments were performed on 2 databases of handwritten digits: CENPARMI and MNIST.

This work is organized as follows: in Section 2 we introduce the 2 wavelets used in the CWT for the preprocessing step: the Mexican hat (in its isotropic and anisotropic versions) and the wavelet gradient. They are respectively based on the second and first derivatives of a Gaussian. In Section 3 we present our proposed feature extraction process, in Section 4 we describe the databases, and in Section 5 we give results and concluding remarks. Contributions and future work are briefly given in Sections 6 and 7.

## 2 2D Continuous Wavelet Transforms

The two-dimensional CWT is the inner product of an image  $s$  with a scaled, rotated and translated version of a wavelet function  $\psi$  [8]:

$$S_{\psi}(b, a, \theta) = a^{-1} \int_{\mathbb{R}^2} \psi(a^{-1} r_{-\theta}(b - x)) s(x) d^2x, \quad (1)$$

where  $x = (x_1, x_2)$ ,  $b = (b_1, b_2) \in \mathbb{R}^2$ ,  $0 \leq \theta \leq 2\pi$ , and

$$r_{\theta}(x) = (x_1 \cos \theta - x_2 \sin \theta, x_1 \sin \theta + x_2 \cos \theta). \quad (2)$$

The wavelet  $\psi$  is highly localized in space; it is either compactly supported or has fast decay. Its integral is zero: for a given scale  $a > 0$  the CWT behaves

like a band-pass filter, providing information on where in the image we can find oscillations or details at that scale. At small scales the CWT captures short-lived variations in color such as thin edges; comparing the CWT at different scales reveals what kind of discontinuity is present; at large scales it blurs the image. If the wavelet is stretched in one direction, the CWT gives information on local orientation in the image.

### 2.1 2D Mexican Hat: Isotropic or Anisotropic

For our wavelet, we choose the Mexican Hat (MH), which is stretched in the direction of one of the axes in accordance with parameter  $\epsilon$  [10]:

$$\psi_{MH}(x_1, x_2) = (2 - (x_1^2 + \frac{x_2^2}{\epsilon})) e^{-(x_1^2 + x_2^2/\epsilon)/2}. \tag{3}$$

Note that when  $\epsilon = 1$ ,  $\psi_{MH}$  is the Laplacian of  $g(x_1, x_2) = e^{-(x_1^2 + x_2^2)/2}$ , a bidimensional Gaussian; it is isotropic, and in that case the CWT gives no information on object orientation. When scaled, its essential support is a disk with radius proportional to the scale. If  $\epsilon \neq 1$ , we have the anisotropic MH, stretched out or shortened, and its support is an ellipse.

The CWT has 4 parameters: scale, angle (orientation), and position  $b = (b_1, b_2)$  in the image. For fixed  $a$  and  $\theta$ , Eq. (1) gives the so-called position representation of the transform. By integrating the CWT's energy over all positions, we have the scale-angle density [6]:

$$E(a, \theta) = \int_{\mathbb{R}^2} |S_\psi(a, \theta, b)|^2 db_1 db_2 \tag{4}$$

### 2.2 The Wavelet Gradient

The first derivatives of  $g$  in each variable give 2 new wavelets:

$$\psi_1(x_1, x_2) = \frac{\partial g(x_1, x_2)}{\partial x_1} \quad \text{and} \quad \psi_2(x_1, x_2) = \frac{\partial g(x_1, x_2)}{\partial x_2}, \tag{5}$$

which have 2 vanishing moments and can be interpreted as a multiresolution differential operator. With both wavelets we construct the wavelet gradient [11] at scale  $a$  and at each position in the image:

$$T_\psi(b, a) = [S_{\psi_1}(b, a, 0), S_{\psi_2}(b, a, 0)]. \tag{6}$$

$S_{\psi_1}(b, a, 0)$ , the first component of  $T_\psi(b, a)$  in Eq. (6), is the horizontal wavelet gradient. Integrating by parts we have

$$S_{\psi_1}(b, a, 0) = -\frac{1}{a} \int_{\mathbb{R}^2} g(a^{-1}(b-x)) \frac{\partial}{\partial x_1} s(x) d^2x, \tag{7}$$

which is equal (except the sign) to the inner product of the first derivative of the image (with respect to  $x_1$ ) and the scaled and translated Gaussian; in other words

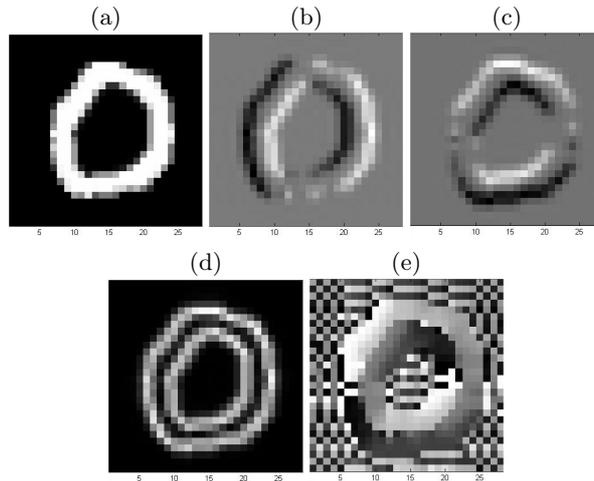
it averages the horizontal gradients smoothed with a Gaussian (with different widths), giving information on vertical edges at various scales. The same holds for the vertical wavelet gradient – the second component  $S_{\psi_2}(b, a, 0)$ , which gives information on horizontal edges at different scales. With the 2 components of the wavelet gradient vector, we may calculate both the absolute values

$$|T_{\psi}(b, a)| = \sqrt{S_{\psi_1}(b, a, 0)^2 + S_{\psi_2}(b, a, 0)^2}, \quad (8)$$

and the angles

$$\angle T_{\psi}(b, a) = \arctan(S_{\psi_2}(b, a, 0), S_{\psi_1}(b, a, 0)) \quad (9)$$

of the transformed digit at each position. We then have the moduli and angular orientation of edges. (The sine and cosine are also calculated, and their signs are analyzed so as to give the angles in interval  $[0 \ 2\pi)$ .) Fig. 1 illustrates the wavelet gradient calculated on a MNIST digit, with  $a = 1$ .



**Fig. 1.** Top row: (a) original digit, (b) horizontal wavelet gradient ( $S_{\psi_1}(b, a, 0)$  in Eq. (6)), (c) vertical wavelet gradient ( $S_{\psi_2}(b, a, 0)$  in Eq. (6)). Bottom row: (d) Absolute values (Eq. (8)), (e) angles (phase) (Eq. (9)) of the wavelet gradient.

In a similar way the second derivatives of the image may be estimated at different scales, and the curvature calculated.

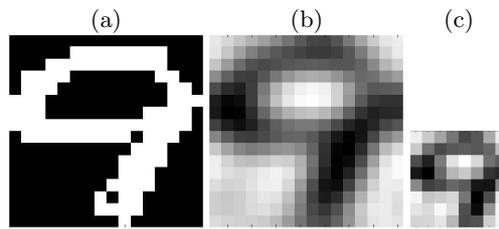
### 3 Preprocessing of Handwritten Digits

#### 3.1 First Step: MH-4 a Small Version with Smooth Edges

A low-level descriptor is extracted from the CWT, that preserves the structure and shape of the image, as well as spatial correlations in all directions: for this

we select the isotropic MH wavelet given in Eq. (3) with ( $\epsilon = 1$ ). By choosing a large scale ( $a = 2.2$ ) to calculate the CWT we obtain a new version of the sample digit, in which the edges are smoothed out and filled in. To reduce dimensionality we subsample the transformed image by 2, by rows and by columns: for this we leave out the odd rows and columns of the transformed image. The result is a smoothed version of the digit, in size a fourth of the original one: we call this descriptor MH-4.

In Fig. 2 we observe an original digit from CENPARMI database, its CWT with isotropic MH, and the obtained small version with smooth edges.



**Fig. 2.** (a) original digit, (b) result of applying the isotropic MH-CWT, (c) the same, subsampled (MH-4)

### 3.2 Second Step: A Complementary Feature Vector

A second high-level descriptor is the complementary feature vector (CFV). This vector has 85 components, averaging information on edges, orientations and curvature at different scales. It is the result of calculating 8 features [A]–[H] over the image, mostly extracted from 2 continuous wavelet transforms: the anisotropic MH-CWT and the gradient wavelet. Each feature is evaluated for different sets of parameters. To select the parameters we chose the ones giving best results, out of a limited set with which we carried out our tests. Each feature and set of parameters produces one scalar value, a component of the CFV.

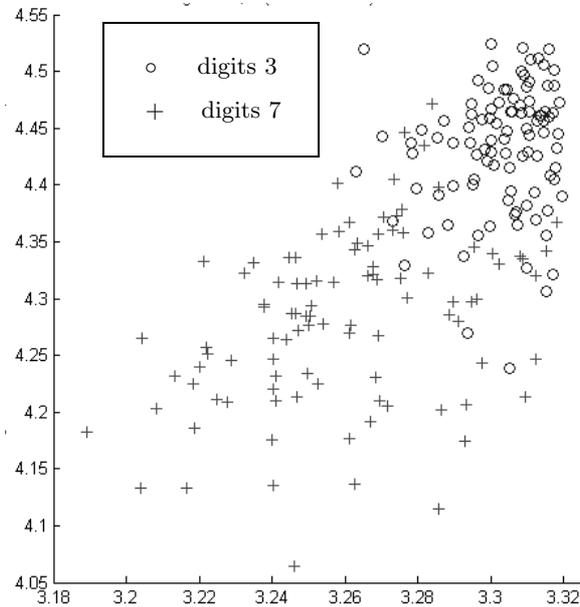
In the case of the MH-CWT (Eqs. (1) and (3))– features [A] and [B]– this scalar value is obtained by means of the sum of squared values (over the transformed digit), which measures the energy, or by means of the entropy, which measures the dispersion of the related histogram.

In the case of the scale-angle density (Eqs. (1), (3) and (4))– feature [C]–  $E(a, \theta)$  is calculated at different scales for a fixed set of angles. The scale giving the greatest energy values is chosen. For that chosen scale  $a$ , the entropy of  $E(a, \theta)$  (varying  $\theta$ ) is calculated.

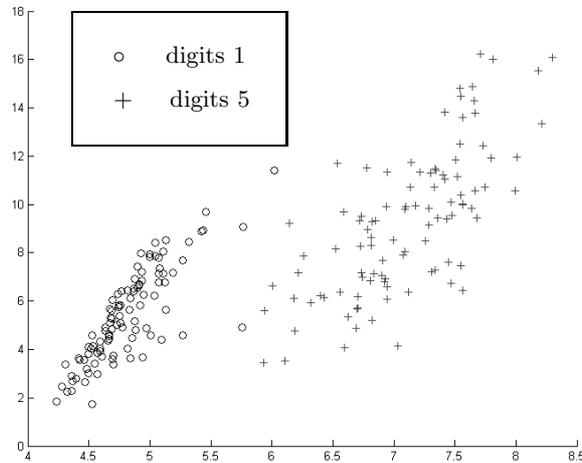
In the case of the wavelet gradient (Eq. (6)), the magnitudes and angles of the transformed digit are found (Eqs. (8), (9)). Either the sum of squared values or the entropy are calculated over the magnitudes – features [D] and [E]. Also the angles are quantized and the entropy of the resulting angles is calculated – feature [F]. The wavelet gradient is also used to calculate the mean curvature – feature [G].

The values of each component are scaled so that their ranges are approximately the same. For features marked with (\*), calculations are performed twice for each set of parameters: first with no thresholding, second with a threshold equal to 30 % of the maximum value. The components of the CFV, grouped by features, are:

- [A] (30 values) Sums of squares of absolute values of the anisotropic MH-CWT, with  $e = 2.5$ , and  $a = 0.8$ , at angles  $\theta = i\pi/10$ ,  $i = 0 \leq i \leq 9$ . The same for  $a = 1$  and  $a = 1.8$ .
- [B] (30 values) Entropy of the absolute values of the anisotropic MH-CWT, with  $e = 2.5$ , and  $a = 0.8$ , at angles  $\theta = i\pi/10$ ,  $i = 0 \leq i \leq 9$ . The same for  $a = 1$  and  $a = 1.8$ .
- [C] (4 values) Entropy of scale-angle density for a fixed scale at angles  $\theta_i = i\pi/N$ ,  $0 \leq i \leq N - 1$ , for  $N = 10$ . The scale-angle density is calculated on the anisotropic MH-CWT with  $e = 2.5$  ( $e = 3.5$ ), at scales 0.8- 3.2 in steps of 0.4, and at angles as mentioned. The scale giving the greatest energy is chosen. The same, for  $N = 15$ .
- [D] (4 values) Sums of squares of absolute values of wavelet gradient for  $a = 1$  and 1.5 (\*).
- [E] (4 values) Entropy of the absolute values of the wavelet gradient, for  $a = 1$  and 1.5 (\*).



**Fig. 3.** Features [F] (y-axis) versus [C] (x-axis) for 200 digits from CENPARMI database



**Fig. 4.** Features [D] (y-axis) versus [E] (x-axis) for 200 digits from MNIST database

[F] (8 values) Entropy of the uniformly quantized angles ( $N$  classes) of the wavelet gradient at a fixed scale, for  $a = 1$  and  $a = 1.5$ , for  $N = 10$  and  $N = 15$  (\*).

[G] (4 values) Curvature calculated on wavelet gradient with  $a=1$ ,  $a = 1.5$  (\*).

[H] (1 value) Curvature calculated on the original image.

In Figs. 3 and 4 we show how 2 components of the CFV (2 features for given parameters) separate 2 classes. The experiment is carried out with 100 digits from 2 distinct classes, and each point in the figures stands for one sample digit, its abscissa and ordinate being the values of the features. Parameters used in Fig. 3: for feature [F] :  $a=1.5$  and 15 angles, for feature [C]:  $e = 2.5$  and 10 angles. Parameters used in Fig. 4: for features [D] and [E],  $a = 1.5$ . Although in the figures the separation is not complete – the separation of classes is achieved with the complementary information of other features – it gives an idea of the efficiency of the features chosen to characterize the classes.

## 4 Databases

Our experiments were performed on the handwritten numeral databases CEN-PARMI and MNIST. These databases have been widely accepted as standard benchmarks to test and compare performances of the methods of pattern recognition and classification. Each database is partitioned into a standard training data set and a test data set such that the results of different algorithms and preprocessing techniques can be fairly compared.

MNIST is a modified version of NIST database and was originally set up by the AT&T group [12]. The normalized image data are available at webpage [13]. MNIST database contains 60,000 and 10,000 graylevel images (of size  $28 \times 28$ ) for training and testing, respectively.

The CENPARMI digit database [14] was released by the Centre for Pattern Recognition and Machine Intelligence at Concordia University (CENPARMI), Canada. It contains unconstrained digits of binary pixels. In this database, 4,000 images (400 samples per class) are specified for training and the remaining 2,000 images (200 samples per class) are for testing. We scaled each digit to fit in a  $16 \times 16$  bounding box such that the aspect ratio of the image would be preserved.

Because the CENPARMI database has less resolution, the database is much smaller, and the sampled digits are less uniform than in the MNIST database, the CENPARMI digits are more difficult to classify.

## 5 Results and Conclusions

In Table 1 the recognition rates of our proposed preprocessing technique on the CENPARMI database are listed, for the training set as well as for testing set. We give the recognition rates of the multilayer perceptron without preprocessing the sample digits. We also compare each step of our preprocessing technique (MH-4 and CFV) separately, and jointly. In the latter case, the smaller smooth image as well as the components of the CFV are fed into the neural network. We give the number of neurons in the first, hidden and final layers considered in the neural network's architecture; the number of neurons in the first layer being the dimension of the input. In Table 2 we have the same results for MNIST.

Notice the importance of preprocessing the digits, reflected by the higher recognition rates of MH-4 over no preprocessing. The CFV on its own, because it gives compact information on edges, orientations, curvature and scales, but lacks information on where they happen, gives lower recognition rates. However,

**Table 1.** Recognition rates for CENPARMI database

Preproc.	Network architecture	% Recog training set	% Recog test set
None	$256 \times 220 \times 10$	99.13	88.95
CFV	$85 \times 170 \times 10$	78.35	70.70
MH-4	$64 \times 150 \times 10$	99.22	91.95
MH-4 + CFV	$149 \times 200 \times 10$	99.22	92.60

**Table 2.** Recognition rates for MNIST database

Preproc.	Network architecture	% Recog training set	% Recog test set
None	$784 \times 110 \times 10$	99.29	97.06
CFV	$85 \times 110 \times 10$	79.88	80.85
MH-4	$196 \times 110 \times 10$	99.46	98.04
MH-4 + CFV	$281 \times 130 \times 10$	99.46	98.22

when added to the MH-4 representation, it improves the recognition rates and gives the best results.

Our proposed preprocessing technique, followed by a general purpose MLP classifier, achieved a recognition rate of 92.60 % on the test set for CENPARMI, and 98.22 % for MNIST. The combination of appropriate features and the reduction in the dimension of the descriptors with which the neural network was trained and tested, improved the performance and the generalization capability of the classifier.

Our results improved upon those mentioned in [3] and [4], the latter giving a recognition rate of 92.20 % for CENPARMI. In [5] the authors give a test error of 1.40 % for MNIST, which is better than ours; here it is difficult to evaluate the performance of our preprocessing method because their preprocessing is followed by a more complex classifier. Recently a classifier (with no preprocessing) was presented, based on the Bhattacharya distance and combined with a kernel approach [15], giving a test error of 1.8% for MNIST; our results are better than theirs. We plan to investigate further the properties of our preprocessing technique in order to reduce the error rate percentage.

## 6 Contributions

We present a novel preprocessing technique for handwritten numerals recognition, that relies on the extraction of multiscale features to characterize the classes. These features are obtained by means of different continuous wavelet transforms, which behave as scale-dependent bandpass filters, and give information on local orientation of the strokes.

The combination of appropriate features and the reduction in the dimension of the descriptors with which a multilayer perceptron neural network was trained and tested, improved the performance and the generalization capability of the classifier, obtaining competitive results over MNIST and CENPARMI databases.

## 7 Future Work

We plan to investigate further the properties of our preprocessing technique in order to improve the accuracy with which the features describe the original digits. We also plan to use more complex classifiers to obtain higher recognition rates. The application of this preprocessing technique to the problem of texture recognition is another goal in our research.

## Acknowledgements

This work has been supported by grants UBACYT X166, X199 and BID 1728/OC-AR-PICT 26001.

The authors wish to thank the anonymous reviewer whose comments helped to improve the quality of this paper.

## References

1. Mallat, S.: A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Analysis Machine Intell.* PAMI-11(7) (1989)
2. de Ves, E., Ruedin, A., Acevedo, D., Benavent, X., Seijas, L.: A new wavelet-based texture descriptor for image retrieval. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) *CAIP 2007*. LNCS, vol. 4673, pp. 895–902. Springer, Heidelberg (2007)
3. Wunsch, P., Laine, A.F.: Wavelet descriptors for multiresolution recognition of handprinted characters. *Pattern Recognition* 28(8), 1237–1249 (1995)
4. Chen, G., Bui, T., Krzyzak, A.: Contour-based handwritten numeral recognition using multiwavelets and neural networks. *Pattern Recognition* 36, 1597–1604 (2003)
5. Bhattacharya, U., Vajda, S., Mallick, A., Chaudhuri, B., Belaid, A.: On the choice of training set, architecture and combination rule of multiple MLP classifiers for multiresolution recognition of handwritten characters. In: *9th IEEE International Workshop on Frontiers in Handwritten Recognition* (2004)
6. Antoine, J.-P., Murenzi, R.: Two-dimensional directional wavelets and the scale-angle representation. *Signal Processing* 52, 256–281 (1996)
7. Antoine, J., Vanderghyest, P., Bouyoucef, K., Murenzi, R.: Target detection and recognition using two-dimensional isotropic and anisotropic wavelets. In: *Automatic Object Recognition V*, SPIE Proc, vol. 2485, pp. 20–31 (1995)
8. Antoine, J.-P., Murenzi, R., Vanderghyest, P.: Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns. *Appl. Comput. Harmon. Anal.* 6, 314–345 (1999)
9. Romero, D., Seijas, L., Ruedin, A.: Directional continuous wavelet transform applied to handwritten numerals recognition using neural networks. *Journal of Computer Science & Technology* 7(1), 66–71 (2007)
10. Kaplan, L.P., Murenzi, R.: Pose estimation of sar imagery using the two dimensional continuous wavelet transform. *Pattern Recognition Letters* 24, 2269–2280 (2003)
11. Jelinek, H.F., Cesar Jr., R.M., Leandro, J.J.G.: Exploring wavelet transforms for morphological differentiation between functionally different cat retinal ganglion cells. *Brain and Mind* 4, 67–90 (2003)
12. LeCun, Y., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278–2324 (1998)
13. LeCun, Y., Cortes, C.: The mnist database of handwritten digits, <http://yann.lecun.com/exdb/mnist/index.html>
14. Suen, C., Nadal, C., Legault, R., Mai, T., Lam, L.: Computer recognition of unconstrained handwritten numerals. *Procs. IEEE* 80(7), 1162–1180 (1992)
15. Wen, Y., Shi, P.: A novel classifier for handwritten numeral recognition. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1321–1324 (2008)